

CHAPTER

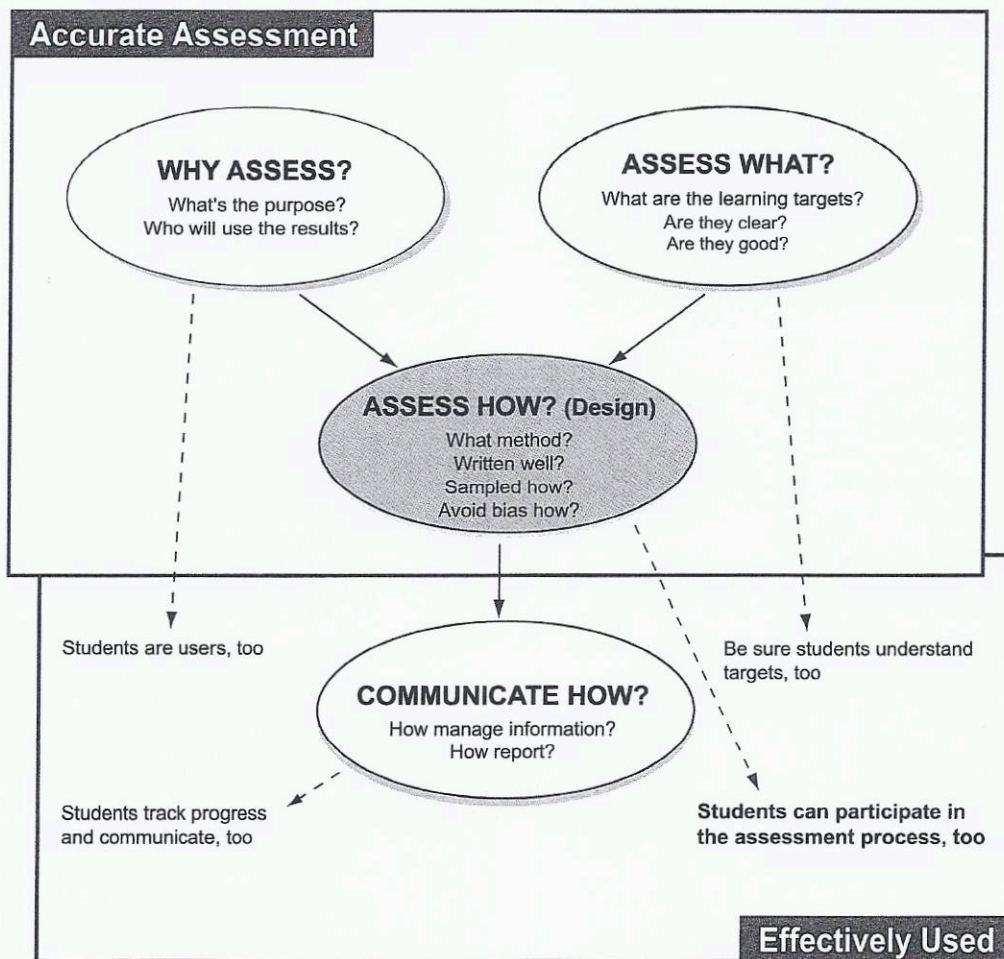
4

Assess How? Designing Assessments to Do What *You* Want

[S]o far, we have examined two keys to assessment quality. The first key is to know at the outset how we intend to use assessment results. Sometimes we can use them to promote learning (assessment *for* learning) and other times to check to see if learning has occurred—that is, for purposes of accountability (assessment *of* learning). As the second key to quality, we have established that assessments must be designed to reflect the variety of achievement targets that underpin standards: mastery of content knowledge, the ability to use knowledge to reason, demonstration of performance skills and product development capabilities. Now we consider the third key to classroom assessment quality—how to design assessments that cover our targets and serve our purposes (the shaded portion of Figure 4.1).

In this chapter we describe four assessment methods representing the range of assessment options, explain how to choose which method to use for any given learning target, and outline the steps in assessment planning and development. We treat each of the four assessment methods in depth in Chapters 5 through 8; here we offer an overview with an emphasis on selecting the proper method and on thoughtful assessment planning.

Figure 4.1 Keys to Quality Classroom Assessment



Assessment Methods—A Menu of Options

Throughout our school careers, both as students and as teachers, we have encountered thousands of different assessments. Although the variations are endless, all of the assessments we have experienced and give today fall into one of four basic categories of methods:

1. Selected response and short answer
2. Extended written response

3. Performance assessment
4. Personal communication

All four methods are legitimate options when their use correlates highly with the learning target and the intended use of the information. (Portions of the following discussion are adapted from Stiggins, 2005.)

Selected Response

Selected response and short answer methods consist of those in which students select the correct or best response from a list provided. Formats include multiple choice, true/false, matching, short answer, and fill-in questions. (Although short answer and fill-in-the-blank do require students to generate an answer, they call for a very brief answer that is counted right or wrong, so we include these options in the selected response category.) For all selected response assessments, students' scores are figured as the number or proportion of questions answered correctly.

Extended Written Response

Extended written response assessment requires students to construct a written answer in response to a question or task rather than to select one from a list. An *extended* written response is one that is at least several sentences in length. Examples include the following:

- Compare pieces of literature, solutions to environmental problems, or economic events.
- Analyze artwork, forms of government, or solutions to problems.
- Interpret music, scientific information, or polling data.
- Solve a mathematics problem and show and explain all work.
- Describe in detail a scientific, mathematical, or economics process or principle, such as how supply and demand works.

We judge correctness of extended written responses by applying one of two types of predetermined scoring criteria. One type gives points for specific pieces of information that are present. For example, when students in a biology class are asked to describe the Krebs cycle, points might be awarded for noting that the cycle describes the sequence of reac-

tions by which cells generate energy, takes place in the mitochondria, consumes oxygen, produces carbon dioxide and water as waste products, and converts ADP to energy-rich ATP. The second type of criteria can take the form of a rubric, such as a general rubric for making comparisons, which can be applied to any exercise calling for comparison.

Scores therefore also take one of two forms: number or percentage of points attained, or rubric scores.

Performance Assessment

Performance assessment is assessment based on observation and judgment; we look at a performance or product and make a judgment as to its quality. Examples include the following:

- Complex performances such as playing a musical instrument, carrying out the steps in a scientific experiment, speaking a foreign language, reading aloud with fluency, repairing an engine, or working productively in a group. In these cases it is the doing—the process—that is important.
- Creating complex products such as a term paper, a lab report, or a work of art. In these cases what counts is not so much the process of creation (although that may be evaluated, too), but the level of quality of the product itself.

As with extended written response assessments, performance assessments have two parts: a performance task or exercise and a scoring guide. Again, the scoring guide can award points for specific features of a performance or product that are present, or it can take the form of a rubric, in which levels of quality are described. For example, to assess the ability to do a simple process, such as threading a sewing machine, doing long division, or safely operating a band saw, points might be awarded for each step done in the correct order. Or, for more complex processes or products, you might have a rubric for judging quality that has several dimensions, such as ideas, organization, voice, word choice, sentence fluency and conventions in writing, or content, organization, presentation, and use of language in an oral presentation. Again, scores could be reported in number or percent of points earned, or in terms of a rubric score.

Personal Communication

Gathering information about students through personal communication is just what it sounds like—we find out what students have learned through interacting with them. Examples include the following:

- Looking at and responding to students' comments in journals and logs
- Asking questions during instruction
- Interviewing students in conferences
- Listening to students as they participate in class
- Giving examinations orally

We usually think of this as informal, rather than formal assessment (in which results are recorded for later use). Often it is. However, as long as the learning target and criteria for judging response quality are clear, information gathered via personal communication can be used to provide descriptive feedback to students, for instructional planning, and for student self-reflection and goal setting. If planned well and recorded systematically, information from personal communication can be used as the basis for assessments *of* learning.

Student responses are evaluated in one of two ways. Sometimes the questions we ask require students to provide a simple, short answer, and all we're looking for is whether the answer is correct or incorrect. This is parallel to scoring for written selected response questions. Questions during instruction usually call for these short answer oral responses.

Other times, student oral responses are longer and more complex, parallel to extended written response questions. Just as with extended written response, we evaluate the quality of oral responses using a rubric or scoring guide. Longer, more complicated responses would occur, for example, during oral examination or oral presentations.

Misconceptions About Assessment Methods

Much “lore” exists about assessment methods, and we address a few of the most common misconceptions here.

1. *Shouldn't we only be using “authentic” assessments—performance assessments—to judge student progress?* None of these methods is inherently superior to any other, and all are viable if used well. Good assessment means clearly knowing what it is you want to assess and then choosing the best method to get the job done, which, as we will show, depends on the purpose and the learning targets being assessed.
2. *I can see how to involve students in assessment when using a performance assessment, but how do you do it with other methods? Doesn't student involvement require performance assessment?* Although many of our strongest examples of student involvement in the past have come from performance assessment applications, there is a vast, untapped reservoir of student-involvement practices leading to higher learning within each assessment method. You will find extended examples in Chapters 5 through 8.
3. *What about portfolios? I notice they aren't listed as a method. Where do they fit in?* Portfolios are a wonderful idea and we devote an entire chapter to their use later in the book. However, they are not an assessment method, but a vehicle for collecting evidence of, tracking, and communicating about student learning. Portfolios offer a way to involve students deeply in the overall process—self-assessment, tracking progress, reflecting on work, goal setting, and communicating about learning. In this sense, portfolios play a valuable role in creating assessment for learning in the classroom.
4. *What about presentations, group projects, worksheets, observations, exhibitions of mastery, posters, and the other ways that teachers gather information?* All of these artifacts and procedures can be classified within the four basic assessment methods described. Presentations and observations are examples of performance assessment. Exhibitions of mastery and group projects can take the form of extended written response, performance assessment, or personal communication depending on how they are carried out. Worksheets are not a method at all because they can contain various types of questions. (Usually worksheets consist of selected response or extended written response questions.) Likewise, posters can be considered either extended written response or performance assessment depending on the assignment and the learning targets being assessed.

Target–Method Match

One of the values in classifying assessments according to method is that we can think clearly about how to assess what we are teaching. The heart of accuracy in classroom assessment revolves around matching different kinds of achievement targets, with all the forms and nuances of each, to the appropriate assessment method. This is easily done and can save time in the long run.

To begin thinking about the match between kind of learning target and assessment method, please complete the following two activities. You may want to discuss possible answers with colleagues.

DEEPEN UNDERSTANDING

Activity 4.1 Which Method?

Let's say you need to assess student achievement on each of the following learning targets. Which assessment method—selected response/short answer, extended written response, performance assessment, or personal communication—would you choose? Please jot down your answers and save them for later reference.

1. Ability to write clearly and coherently
2. Group discussion proficiency
3. Reading comprehension
4. Proficiency using specified mathematical procedures
5. Proficiency conducting investigations in science

Activity 4.2 Target–Method Match

For this activity, you will determine which assessment method is the best match for each of the four kinds of learning targets: knowledge, reasoning, skill, and product targets. To do this you will need to read through the following four scenarios and record your answer to each question by marking an “X” in the appropriate box on Figure 4.2 when your answer is “Yes.” (A printable version of the figure is on the accompanying CD in the file, “Target–Method Match Chart.”) You can put an “X” in more than one box. You can use capital “X” to denote really good matches, and a checkmark to denote an acceptable match under certain conditions (or whatever easily distinguishable marks you wish). On a separate sheet of paper, write your justifications for each answer, as requested. To make your choices, think about accuracy and efficiency: which methods will provide the most accurate information with the highest degree of efficiency? If you are working with a learning team, consider discussing your responses as a group.

Scenario 1: Assessing Student Mastery of Content Knowledge

Scenario: You want your students to master specific subject matter knowledge because it represents an important foundation for later work. You plan a series of instructional activities to help your students reach this goal. Now you want to assess to be sure they’ve got it. In this particular case, you want them to know the material outright, not through the use of reference materials.

Question 1: Should you assess mastery of this material using selected response or short answer modes of assessment, such as multiple choice, true/false, or matching exercises? Briefly explain your response.

Question 2: Should you assess your students’ mastery of this material using an extended written response form of assessment? Defend your answer.

Question 3: Should you use a performance assessment to assess students’ mastery of this content knowledge? Defend your answer.

Question 4: Do you think the personal oral communication form of assessment—by oral exam, interview, conference, or discussion—could viably assess your students’ mastery of this content knowledge? Why or why not?

Activity 4.2 (Continued)

Scenario 2: Assessing Reasoning Proficiency

Scenario: You are a teacher who has seen to it that your students are able to access important knowledge when required. Now you want to see if they can use that knowledge productively to solve relevant problems. You want to see if they can reason analytically (think about the parts of things) and comparatively (think in terms of similarities and differences), draw inferences, and think critically (take and defend a position on an issue, for example).

Question 1: Can you get at these things with selected response or short answer assessments? Why or why not?

Question 2: Does extended written response assessment work in contexts where we seek to assess reasoning proficiency? Why or why not?

Question 3: Is performance assessment a viable alternative? Why or why not?

Question 4: Can we use personal oral communication as an assessment method to probe a student's ability to reason effectively and solve problems? Defend your response.

Scenario 3: Assessing Mastery of Skills

Scenario: You teach French and wish to assess your students' skill at communicating in that language in a conversational situation. So the skill of *oral language proficiency* is your target.

Question 1: Can you assess oral language proficiency in a conversational context using a selected response or short answer mode of assessment? Defend your answer.

Question 2: Can you assess these skills using extended written response assessment? Why or why not?

Question 3: Will performance assessment work as a basis for assessing the foreign language speaking proficiency of your students? Why or why not?

Question 4: Can you use personal oral communication as a basis for assessing conversational skill in a second language? Defend your response.

Question 5: Would your responses also apply to other skills such as operating a sewing machine, dribbling a basketball, or reading aloud fluently?

Activity 4.2 (Continued)

Scenario 4: Assessing the Ability to Create Quality Products

Scenario: You want your students to be able to create quality products—products that meet certain specified standards. They might be samples of writing, term papers, technology products, craft products, artistic creations, or others. Your instruction has centered on helping students learn the differences between products that are of high and low quality. You have provided practice in developing products that meet your standards. Now it is time to assess the students' achievement to see if your instruction was effective.

Question 1: Can you assess the ability to create these kinds of products using selected response or short answer modes of assessment? Why or why not?

Question 2: Will extended written response assessment work for evaluating this kind of achievement? Explain.

Question 3: Can performance assessment provide the evidence of proficiency needed to evaluate this kind of achievement target? Defend your response.

Question 4: Is personal oral communication a viable way to assess when creation of a product is the target? Why or why not?

Checking Your Matches

Table 4.1 identifies the strong matches between kinds of achievement targets and assessment methods. Please compare the table to the target-method match choices made by your team. Note and discuss discrepancies.

Figure 4.2 A Plan for Matching Assessment Methods with Achievement Targets

Target to Be Assessed	Assessment Method			
	Selected Response	Extended Written Response	Performance Assessment	Personal Communication
Knowledge Mastery				
Reasoning Proficiency				
Performance Skills				
Ability to Create Products				

Source: Adapted from *Student-Involved Assessment for Learning*, 4th ed. (p. 65), by R. J. Stiggins, 2005, Upper Saddle River, NJ: Merrill/Prentice Hall. Copyright © 2005 by Pearson Education, Inc. Adapted by permission of Pearson Education, Inc.

Assessing Knowledge Targets

Selected Response

There is usually no argument about this match. Selected response options do a good job at assessing mastery of discrete elements of knowledge, such as important history facts, spelling words, foreign language vocabulary, and parts of plants. These assessments are efficient in that we can administer large numbers of questions per unit of testing time and so can cover a lot of material relatively quickly. Thus, it is easy to obtain a good sample of student knowledge so that we may infer level of overall knowledge acquisition from the sample on the test.

Table 4.1 Links Among Achievement Targets and Assessment Methods

Target to Be Assessed	Assessment Method			
	Selected response	Extended Written Response	Performance Assessment	Personal Communication
Knowledge Mastery	Good match for assessing mastery of elements of knowledge.	Good match for tapping understanding of relationships among elements of knowledge.	Not a good match—too time consuming to cover everything.	Can ask questions, evaluate answers and infer mastery—but a time-consuming option.
Reasoning Proficiency	Good match only for assessing understanding of some patterns of reasoning.	Written descriptions of complex problem solutions can provide a window into reasoning proficiency.	Can watch students solve some problems and infer reasoning proficiency.	Can ask student to “think aloud” or can ask followup questions to probe reasoning.
Skills	Not a good match. Can assess mastery of the knowledge prerequisites to skillful performance, but cannot rely on these to tap the skill itself.		Good match. Can observe and evaluate skills as they are being performed.	Strong match when skill is oral communication proficiency; not a good match otherwise.
Ability to Create Products	Not a good match. Can assess mastery of knowledge prerequisite to the ability to create quality products, but cannot use to assess the quality of products themselves.	Strong match when the product is written. Not a good match when the product is not written.	Good match. Can assess the attributes of the product itself.	Not a good match.

Source: Adapted from *Student-Involved Assessment for Learning*, 4th ed. (p. 69), by R. J. Stiggins, 2005, Upper Saddle River, NJ: Merrill/Prentice Hall. Copyright © 2005 by Pearson Education, Inc. Adapted by permission of Pearson Education, Inc.

Extended Written Response

Extended written response is useful for assessing blocks of knowledge rather than pieces of information detached from one another—causes of environmental disasters, the carbon cycle in the atmosphere, how one mathematical formula can be derived from another, or the concept of checks and balances in government. Extended written response assessments are not as efficient as selected response assessments because responding to each takes longer, but they can get at deeper levels of knowledge.

Performance Assessment

Performance assessment is usually not a good choice for assessing knowledge targets, for three reasons.

We'll illustrate the first reason with a brief example. Let's say we ask a student to complete a rather complex performance, such as writing and executing a computer program, for the purpose of determining if she has the prerequisite knowledge. If the student successfully executes the program, then we know that she possesses the prerequisite knowledge. The problem comes in when the program does not run successfully. Was it due to lack of knowledge of the programming language, due to the inability to use knowledge to create a program that does what it is intended to do, or merely due to the inability to manipulate the keyboard or to proofread? We can't know the reason for failure unless we follow up the performance assessment with one of the other assessment methods. We must ask some short answer or extended response questions to find out if the prerequisite knowledge was there to start with. But, if our initial objective was to assess mastery of specific knowledge, why go through the extra work? To save time and increase accuracy, we recommend using selected response, short answer, and extended written response assessments to evaluate knowledge targets.

The second reason this is not a good match is because it is inefficient to assess all content knowledge with a performance assessment. A single performance task does require some subset of knowledge, and you can assess its presence with a particular performance task, but how many performance tasks would you have to create to cover all the knowledge you want students to acquire? For example, how many performance assessments would it take to determine if students can spell all the words you want them to spell? Or, how many performance assessments would it take to determine if students can perform all the mathematical operations they have been taught in a semester? Again, we recommend assessing knowledge with a simpler method and reserving performance assessment for those learning targets that really require it.

The third reason that performance assessments are usually not a good match for knowledge learning targets has again to do with practicality. It just isn't practical (or safe) to conduct some performance assessments. For example, let's say that you want to know if students can read schedules, such as bus schedules. It would be most "authentic" to ask students to get around town on the bus, but it would be highly inefficient and perhaps dangerous. Asking students to answer multiple-choice or short answer questions requiring understanding of a bus schedule would be a good compromise for getting the information needed.

Personal Communication

This is a good match with knowledge targets for most students at all grade levels, but tends to be inefficient if a lot of knowledge is to be assessed for lots of students. Personal communication works best for real-time sampling of student understanding during instruction. Also, for some students, such as those with special needs, English language learners, or younger students, it is the best way to gather accurate information.

Assessing Reasoning Proficiency

Selected Response

A common misunderstanding is that selected response questions can't assess reasoning proficiency. Although not a good choice for some patterns of reasoning, other patterns of reasoning can be assessed in selected response format. For example:

- Which of the following statements best describes how dogs in real life are different from the dog in the story? (Comparative reasoning)
- What generalization can you make from this selection about how these plants lure their prey? (Inference—generalizing)
- Which answer best explains the author's purpose in writing this story? (Inference—determining author's purpose)
- Choose the sentence that best tells what the story is about. (Inference—identifying main idea)

There are limits to this format when assessing reasoning. If you want to assess how well students can choose from their store of reasoning proficiencies to solve a problem, solve a problem requiring several steps, explain their choice or reasoning process, or defend an opinion, you must use another assessment method. For example, you might ask students

to solve the following problem in mathematics: “Estimate the number of hours of TV advertising the typical U.S. fifth grader watches in a year. Describe your procedure for determining your answer.” This is an extended response question. If the learning target you want to assess is student reasoning, a single number as the right answer is not the focus of the assessment—the process itself is.

Extended Written Response

Extended written response is a good choice for assessing reasoning targets. Students can be encouraged to write to examine their own thinking. The trick here is to write good questions, ones that require students to analyze, compare, contrast, synthesize, draw inferences, and evaluate *novel* information. For example, if you want students to be able to determine the type of government present in a country (such as democracy, theocracy, dictatorship, or monarchy), you could teach the characteristics of each and practice identifying them. Then, on the assessment, you would give the characteristics of a particular government in a country they had not studied and ask students to identify its type and to explain their reasons.

Remember that to assess reasoning, the question has to be novel. If students worked on the answer to the question during instruction, then the answer is a piece of remembered knowledge, which does not require reasoning. For example, consider the following science standard: “Recognize and analyze alternative explanations and models” (National Research Council, 1996, p. 175). The intent of this standard is that students will be able to use scientific criteria to determine the most plausible explanation or model. To assess this aspect of scientific inquiry, students must be asked to evaluate an explanation or model different than the ones the teacher has used for demonstration and also different from the ones the students have practiced with.

Performance Assessment

This is a partial match for assessing reasoning. For example, we can observe students carrying out science laboratory procedures and draw conclusions about their reasoning based on our observations. But, there’s a hitch that keeps performance assessment from being a great match with reasoning targets: we need to make an inference from what we observe. If students do well on a performance task requiring specific patterns of reasoning, we can assume that reasoning is sound. However, if they don’t do well, it could be due to lack of prerequisite knowledge, lack of motivation, or to imprecise reasoning. Without engaging in additional time-consuming assessment, we may not be able to judge level of achievement on reasoning targets.

Personal Communication

For gathering accurate information, personal communication is a strong match to reasoning targets. Teachers can ask students questions to probe more deeply into a response. Or, students can demonstrate their solution to a problem, explaining their reasoning out loud as they go.

The drawbacks with using personal communication to assess reasoning proficiency are, as always, the amount of time it takes and the record-keeping challenge it poses.

Assessing Performance Skills Targets

There is really only one assessment method that adequately covers performance skills targets, and that is performance assessment. We can use other assessment methods to determine if students possess the knowledge required to perform skillfully, but the only way to determine whether students can actually perform skillfully is to watch them do it and then judge their level of achievement. For example, we can ask students to answer selected response or oral questions about how to conduct themselves during a job interview, but the only way to determine how well they can do it is to watch them during a simulated job interview.

Performance assessment overlaps with personal communication when the performance skills in question fall into the category of oral proficiency, such as speaking a foreign language or giving an oral presentation.

Assessing Proficiency in Creating Products

As with performance skills, the only option for determining whether students can create a certain kind of product is performance assessment: have them create the product or performance and then judge its quality. Once again, we can assess the knowledge required for creating a quality product with a less time-consuming method, but the only way to determine students' levels of proficiency in creating the product is to have them create it.

Performance assessment overlaps with extended written response when the product in question requires writing, such as writing a business letter, lab report, research report, or health and fitness plan.